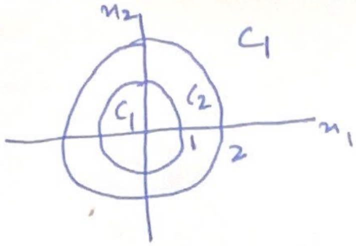
Harshit Gupta
2019ME20885

Q1 a.)



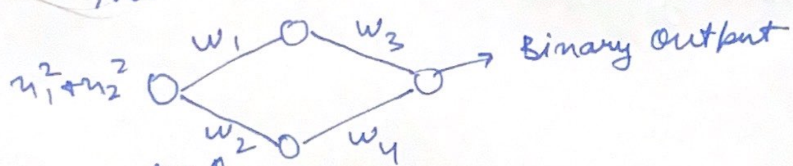
we can make a new feature as $x_1^2+x_2^2$

~~and the neural network can be as~~

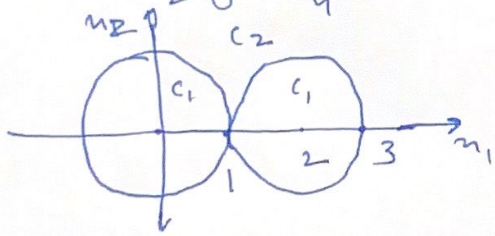
~~tag~~ as ~~these~~ class C_2 can by recognized.

by decision boundary $1 < x_1^2 + x_2 < 2$ and all the other points will be taken in class C_1.

Neural Network



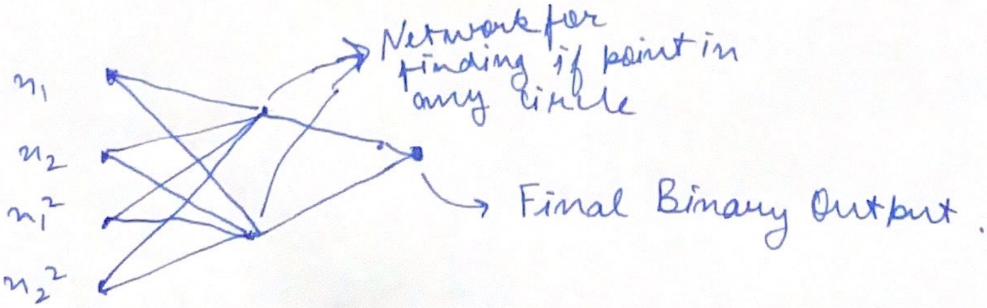
b)



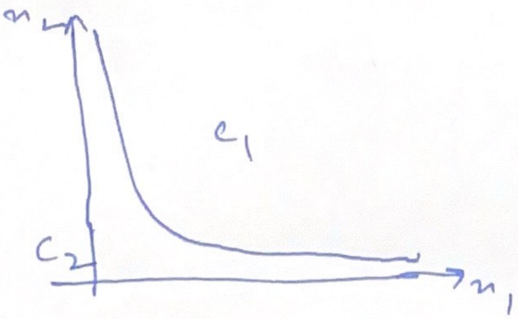
$(x_1-2)^2 + x_2^2 \leq 1$

$x_1^2 + x_2^2 \leq 1$

We can make a similar model to get a circular boundary and Engineer new feature like x_1^2, x_2^2, $x_1^2+x_2^2$. To get whether a point lies in any of the circle and then use ~~get~~ "OR" to get the final output

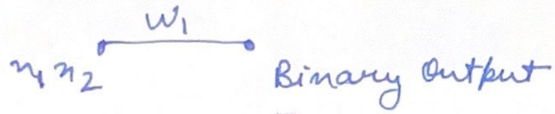


c.)



For this problem we can add a feature of $x_1 x_2$ as there is decision boundary at $x_1 x_2 = 1$ and $x_1 x_2 > 1$ goes to C_1 whereas $x_1 x_2 < 1$ goes to C_2.

Neural Network



Harshit Gupta
2019 ME20885

Q2 $X, Y \sim N(\mu, \Sigma)$

$X \in \mathbb{R}^{m}$

$\mu \in \mathbb{R}^{m+1}$

$\Sigma \in \mathbb{R}^{m+1 \times m+1}$

$$P(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{m/2} \times |\Sigma|^{1/2}} \exp\left(\frac{-1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

Using Bayes Theorem

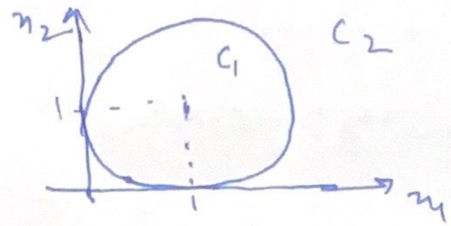
$$P(Y \mid X) = \frac{P(X \mid Y)\, P(Y)}{P(X)}$$

$$y' = N(x-\mu, \Sigma)$$

Harshit Gupta
2019ME20885

$(x_1-1)^2 + (x_2-1)^2 \leq 1$

Q3



a.) Logistic Regression if used directly for x_1 and x_2 the model won't work well ~~[illegible]~~.
Because Logistic Regression requires data to be linearly seperable to make good decision boundaries.

b.) As the data is not linearly seperable directly, we need to project it to a higher feature dimention so we will add new features like x_1, x_2, x_1^2, x_2^2 ~~[illegible]~~
because using these features we can create a equation of circle which is the decision boundary for our model.

c.) Log likelihood for classification problems is given

$$\log L(X,Y;W) = \log \prod_{i=1}^{n} (\hat{y}_i)^{y_i} (1-\hat{y}_i)^{1-y_i}$$

$$= \sum_{i=1}^{n} (1-y_i) \log (1-\hat{y}_i) + y_i \log \hat{y}_i$$

where $\hat{y}_i = \sigma(w^T x_i)$ ↳ sigmoid function

$$\sigma(w^T x_i) = \frac{1}{1+e^{-w^T x_i}}$$

$+e^{-z}$
$+2(1+e^{-z})$
$\frac{1}{2(e^{z}+1)}$

d.) $\log(L(X,Y;W)) = \sum (1-y_i)\log(1-\hat{y}_i) + y_i \log(\hat{y}_i)$

$\frac{d\,\sigma(x)}{dx} = \sigma(x)(1-\sigma(x))$

$= (1-y_i)\log(1-\sigma(W^T x)) + y_i \log(\sigma(w^T x))$

$$\frac{\partial \log L}{\partial w_j} = \sum \frac{(1-y_i)}{1-\sigma(w^T x)} \times -1\, \sigma(w^T x)(1-\sigma(w^T x)) \times x_{ij} + \frac{y_i}{\sigma(w^T x)} \times \sigma(w^T x)(1-\sigma(w^T x)) \times x_{ij}$$

$$= \sum (y_i - 1)\,\sigma(w^T x)\, x_{ij} + y_i (1-\sigma(w^T x))\, x_{ij}$$

$$= \sum_{i=1}^{N} y_i x_{ij} - \sigma(w^T x)\, x_{ij}$$

$$\frac{\partial \log L}{\partial w_j} = \sum_{i=1}^{M} (y_i - \sigma(w^T x_i))(x_{ij})$$

$$\boxed{\frac{\partial \log L}{\partial W} = X^T[Y - \hat{Y}]}$$